



**Meeting/Teleconference Name, Date**

**Name of Meeting**

|                                   |   |  |  |
|-----------------------------------|---|--|--|
| <b>Date, Time &amp; Location:</b> | October 18, 2004 2:00-3:00 EDT  |  |  |
| <b>Attendees:</b>                 | <i>Attendee, Cancer Center, etc.</i><br><br>Sinoula      Apostolou      Fox Chase Cancer Center<br>Steven      Eschrich      Moffitt Cancer Center<br>Zhangzhi      Hu      Georgetown Medical Center<br>Michael      Keller      Booz Allen Hamilton<br>Juli      Klemm      3rd Millennium<br>Simon      Lin      Duke University<br>James      Lyons-Weiler      Pittsburgh<br>Patrick      McConnell      Duke University<br>Thomas      Moloshok      Fox Chase Cancer Center<br>Salvatore      Mungal      Duke University<br>Michael      Ochs      Fox Chase Cancer Center<br>Scott      Oster      Ohio State Univ<br>Joshua      Philipps      NCICB<br>Jarek      Puszynski      SAIC<br>Liat      Shimoni      Fox Chase Cancer Center<br>Jennifer      Shoemaker      Duke University Medical Center<br>Craig      Street      Penn<br>Qing      Xiao      JPL |  |  |
| <b>Introduction:</b>              | <b><u>Roll call, open meeting, review meeting goals</u></b><br><br>- Patrick McConnell and Simon Lin (Duke) will give presentation on MIAPE and mzXML<br>- Introduction of new proteomics SIG lead – Sinoula Apostolou<br><br><b><u>Announcements</u></b><br><br>- Appointed facilitators for SIG<br>– Leo Cheung from U. Hawaii. Will help on BCDE from cross country workspace.<br>Work on beta standards, controlled vocabularies<br><br>- Meeting will be held for new facilitators in ~ 2 weeks  |  |  |
| <b>Follow up on action items:</b> | <b><u>Follow up on last meeting's action items</u></b><br><br>Duke and FCCC sharing out data models for SIG<br><br>- FCCC has most of the documents together. Will wait until the next meeting where LIMS is discussed.   |  |  |



**Meeting/Teleconference Name, Date**

|               |   |
|---------------|---|
|               | <ul style="list-style-type: none"><li>- FCCC have comparison of structure between MIAPE standard</li><li>- Duke is at the stage where the white papers are being drafted. Do not have solid data models. May have data model by next meeting.</li></ul>   |
| Presentation: | <p><u>mzXML presentation by Patrick McConnell and Simon Lin</u></p> <p>Slides can be found at the following link:<br/><a href="http://cabig.nci.nih.gov/workspaces/ICR/Meetings/SIGs/Proteomics/20041018_mzXML_presentation">http://cabig.nci.nih.gov/workspaces/ICR/Meetings/SIGs/Proteomics/20041018_mzXML_presentation</a></p> <p>The link to information about mzXML is:<br/><a href="http://sashimi.sourceforge.net/software_glossolalia.html">http://sashimi.sourceforge.net/software_glossolalia.html</a></p> <p><u>Open discussion</u></p> <p>The group discussed mzXML</p> <ul style="list-style-type: none"><li>- MzXML is intended for proteomics data I/O. However, the general utility of XML breaks down with large data sets. It may be possible to address this issue with Base64 encoding, for example.</li><li>- There is overlap between the MIAPE and the mzXML efforts. It is likely that PSI-ML will be superseded by mzXML</li><li>- The concept of "study" is not represented in mzXML. Thus, database developers and data producers should work together to determine which data elements are required to produce a study. This requirement should be fed back to the developers of mzXML.</li><li>- Note that XML is not a database or an object model, though one can write it to implement an object model.</li><li>- mzXML, along with the MIAPE efforts, represents an immature, emerging standard. To deal with this, one should design software with a flexible architecture and have the ability to migrate to the standards as they evolve. Build what you can today, but don't lock into anything too specific. Standards take a long time!</li></ul> <p>For raw data storage</p> <ul style="list-style-type: none"><li>- MIAPE describes minimum data elements for proteomics experiments</li><li>- PSI-ML describes data elements however, is still abstract</li></ul> <p>Standards</p> <p>The MIAPE effort has two arms: 1) PSI-ontology, which will ultimately be a part of the MGED ontology and 2) PSI-ML, which describes the actual data elements. Pedro was a precursor to MIAPDuke is collaborating with the EDNR to look into merging their schemas. <u>Zhanghi Hu from the Protein Information Resource (PIR) at Georgetown University Medical Center met</u> with Rolf Apweiler on October 19. Rolf is the head of EBI Sequence Database Group and of the HUPO PSI group. He presented some updated information on the current state of MIAPE and standard data formats (PSI-ML, mzData and mzIdent), which will be presented by Rolf in the HUPO 3rd ANNUAL WORLD CONGRESS in Beijing.</p> |



**Meeting/Teleconference Name, Date**

There are two modules in the PSI-ML, mzData and mzIdent. mzData is used for mass spectrometry data and is "now in a mature state", and mzIdent is for a 'search engine' (peptide mapping) and is ongoing. There were comparisons between mzXML and PSI-ML in the SIG discussion. It seemed that the audiences' impression was that PSI-ML is still abstract, while mzXML is more tangible. There seemed to be a consensus emerged from the group that PSI-ML will be superceded by mzXML. My question is, given the current state of the PSI-ML, what would be the best way to adopt a standard format for proteomics data, one or the other, or a merged/combined version of the two? I would be interested in knowing your opinions as well as of others in the group. I would also be glad to get more information from the HUPO PSI group if necessary.

**Ontologies and Controlled Vocabularies**

- discussion on time frame for ontology. May take many years. JPL is exploring an ontology-driven model for core data elements. The advantage to this approach is that while terms may change, an ontology provides the concept of relationships and can potentially be merged with other ontologies.
- The PSI-ontology effort is at a very early stage. This group may need to seek the help of the VCDE group toward the issue of ontologies and controlled vocabularies for proteomics

The group agreed that mzXML should be adopted as the standard of data exchange within the Proteomics SIG.

-

**Action  
Items:**

| Name Responsible  | Action Item  | Date Due | Notes |
|-------------------|--|----------|-------|
| Michael Ochs      | Share current resources to SIG   | 11/1/04  |       |
| Qing Xiao         | Send a summary of the requirements of the SIG with respect to ontology                             | 11/1/04  |       |
| Patrick McConnell | Email slides and link to XML   | 10/18/04 |       |
| Juli Klemm        | Contact the VCDE workspace about the decision of this SIG to use mzXML as a data exchange standard | 11/1/04  |       |